Censored Linear Regression for Case-Cohort Studies

Bin Nan, Menggang Yu, and John D. Kalbfleisch

University of Michigan, Indiana University, and University of Michigan

October 12, 2004

Abstract

Right censored data from a classical case-cohort design and a stratified case-cohort design are considered. In the classical case-cohort design, the subcohort is obtained as a simple random sample of the entire cohort, whereas in the stratified design, the subcohort is selected by independent Bernoulli sampling with arbitrary selection probabilities. For each design and under a linear regression model, methods for estimating the regression parameters are proposed and analyzed. These methods are derived by modifying the linear ranks tests and estimating equations that arise from full-cohort data using methods that are similar to the "pseudo-likelihood" estimating equation that has been used in relative risk regression for these models. The estimates so obtained are shown to be consistent and asymptotically normal. Variance estimation and numerical illustrations are also provided.

KEY WORDS: Case-cohort design; Censored linear regression; Counting processes; Martingales; Rank statistic.

1 Introduction

Since first introduced by Prentice (1986), Case-cohort designs have been widely used in large epidemiologic cohort studies and disease prevention trials. In a case-cohort design, the complete covariate information is processed only for the failures and a random sample of the entire cohort, the so called subcohort. We call the design a classical case-cohort design if the subcohort is a simple random sample of the cohort, or a stratified case-cohort design if the subcohort is a stratified sample of the cohort. Prentice (1986) proposed pseudo-likelihood estimators for relative risk parameters of a Cox model in a classical case-cohort design. Large sample properties were then studied in the follow-up paper by Self and Prentice (1988). Kulich and Lin (2000) proposed an additive hazards model for case-cohort studies that allows estimation absolute risk parameters.

In contrast to modelling hazard functions, a linear regression model postulates a direct relationship between the respone and the covariates with a corresponding parameter interpretation. It thus becomes an important alternative for analyzing censored survival data. If the complete covariate information is collected for the entire cohort, then coefficients in a linear model with right censored data can be estimated using a linear rank statistic. Large sample properties of the corresponding estimator have been studied by Tsiatis (1990), Ritov (1990) and Ying (1993). In this article, we propose estimators for a classical case-cohort design and a stratified case-cohort design. Large sample properties of these estimators are studied in a manner similar to the work of Tsiatis (1990) and Self and Prentice (1988).

In Section 2, we introduce the linear regression model for case-cohort studies. Sections 3 and 4 respectively outline asymptotic properties of the classical case-cohort design and the stratified case-cohort design. In Section 5, we propose an easily computed variance estimator. We give numerical illustrations in Section 6 and a brief discussion in Section 7. An appendix contains most of the proofs.

2 The Censored Linear Model

Let T and C be monotonically transformed failure and censoring times obtained from a known transformation. The log transformation is often used in practice to give the accelerated failure time model (e.g. Kalbfleisch and Prentice, 2002). For subject i in the cohort, we observe $X_i \equiv T_i \wedge C_i$ and the failure indicator $\Delta_i \equiv I\{T_i \leq C_i\}$. Let Z_i be the d-dimensional covariate. For notational simplicity we assume $d = 1$, although results in this article should hold for $d > 1$ as discussed in Tsiatis (1990). The model is

$$T_i = \beta_0 Z_i + e_i, \quad i = 1, \ldots, n,$$

where n is the total number of individuals in the cohort, and given (Z_i, C_i) the e_i's are independent and identically distributed with an unknown distribution.

When (Z_i, X_i, Δ_i) are observed for the entire cohort, Tsiatis (1990) introduced the estimating function for β_0,

$$S_n(\omega_n, \beta_0) = \sum_{i=1}^{n} \int \omega_n(u, \beta_0)\{Z_i - \bar{Z}(u, \beta_0)\}\ dN_i(u + \beta_0 Z_i) \tag{1}$$

where $\bar{Z}(u, \beta_0) \equiv D^{(1)}(u, \beta_0) \Big/ D^{(0)}(u, \beta_0)$,

$$D^{(1)}(u, \beta_0) = n^{-1} \sum_{j=1}^{n} Z_j Y_j(u + \beta_0 Z_j) \quad \text{and} \quad D^{(0)}(u, \beta_0) = n^{-1} \sum_{j=1}^{n} Y_j(u + \beta_0 Z_j),$$

$\omega_n(u, \beta_0)$ is a weight process, and $N_i(u + \beta_0 Z_i)$ is the failure counting process for subject i. Very often $\omega_n(u, \beta_0)$ is chosen to be 1, which corresponds to the log-rank test. The right hand side of (1) is very similar to the estimating function for a Cox regression, and as for the Cox model, martingale theory can be used to investigate large sample properties of $S_n(\omega_n, \beta_0)$.

Let $\lambda(\cdot)$ be the hazard function of e_i. It is easily seen that $\sum_{i=1}^{n} Y_i(u + \beta_0 Z_i)\{Z_i - \bar{Z}(u, \beta_0)\} = 0$ and hence

$$S_n(\omega_n, \beta_0) = \sum_{i=1}^{n} \int \omega_n(u, \beta_0)\{Z_i - \bar{Z}(u, \beta_0)\}\ dM_i(u + \beta_0 Z_i)\ ,$$

where

$$M_i(u + \beta_0 Z_i) = N_i(u + \beta_0 Z_i) - \int_{-\infty}^{u} Y_i(v + \beta_0 Z_i)\lambda(v)\ dv$$

is a martingale process with respect to the filtration

$$\mathcal{F}_n(u, \beta_0) \equiv \sigma\Big[I\{X_i - \beta_0 Z_i \leq u\}, \Delta_i I\{X_i - \beta_0 Z_i \leq u\}, Z_i;\ i = 1, \cdots, n\Big]\ .$$

The complication, however, is that $S_n(\omega_n, \beta)$ in (1) is a step function of β instead of a continuous function as in the case of the Cox regression. Tsiatis (1990) showed that

$S_n(\omega_n, \beta)$ is asymptotically linear in a $n^{-1/2}$-neighborhood of the true value β_0, the true value of β. Using this result, the proof of the asymptotic normality of a zero-crossing of $S_n(\omega_n, \beta)$ became straightforward.

In a case-cohort study, we observe complete data (Z_i, X_i, Δ_i) only when subject i is an observed failure or a member of the subcohort. Let $\mathcal{D}$ denote the set of failures observed during the study period, and $\mathcal{C}$ denote the subcohort. Note that the intersection of these two sets may not be empty. For a classical case-cohort design, $\mathcal{C}$ is a simple random sample of size $\tilde{n}$ from the entire cohort. Similar to the approach of Self and Prentice (1988) for the Cox model, we propose the estimating function

$$\widetilde{S}_n(\beta_0) = \sum_{i=1}^{n} \int \{Z_i - \widetilde{Z}(u, \beta_0)\}\ dN_i(u + \beta_0 Z_i)\ , \tag{2}$$

where $\widetilde{Z}(u, \beta_0) = \widetilde{D}^{(1)}(u, \beta_0) \Big/ \widetilde{D}^{(0)}(u, \beta_0)$ with

$$\widetilde{D}^{(\ell)}(u, \beta_0) = \tilde{n}^{-1} \sum_{j \in \mathcal{C}} Z_j^{\ell} Y_j(u + \beta_0 Z_j), \quad \ell = 0, 1, 2.$$

Note that $D^{(2)}$ is used later. Note also that we have taken $\omega_n(u, \beta_0) = 1$ in (2) for simplicity. The results should hold for the weight functions discussed in Tsiatis (1990). Note that in (1), $\bar{Z}(u, \beta_0)$ is calculated using full-cohort data, whereas in (2), $\widetilde{Z}(u, \beta_0)$ is calculated using the subcohort data only.

If a correlate of Z, say Z^*, is available for all the subjects in the cohort, available literature suggests that selecting the subcohort using stratified sampling based Z^* can improve efficiency in hazard regression models. We expect a similar result should hold for censored linear models and our simulations have supported this. There are many sampling schemes for selecting a stratified subcohort. In this article, we consider an independent Bernoulli sampling method where $P(i \in \mathcal{C} | Z_i^*) = \pi(Z_i^*)$, $i = 1, \ldots, n$ independently.

Let $R_i = I(i \in \mathcal{C})$, and $W_i(Z_i^*) = R_i / \pi(Z_i^*)$ be the weight for subject i. We assume $\pi(Z_i^*) > \rho > 0$ for all i and some constant ρ to ensure bounded weights. For the stratified case-cohort study, we propose the estimating function,

$$\widetilde{S}_n^B(\beta_0) = \sum_{i=1}^{n} \int \{Z_i - \widetilde{Z}^B(u, \beta_0)\}\ dN_i(u + \beta_0 Z_i)\ , \tag{3}$$

where $\widetilde{Z}^B(u,\beta_0) = \widetilde{D}_B^{(1)}(u,\beta_0) \Big/ \widetilde{D}_B^{(0)}(u,\beta_0)$ with

$$\widetilde{D}_B^{(\ell)}(u,\beta_0) = n^{-1}\sum_{j=1}^{n} W_j Z_j^{\ell} Y_j(u+\beta_0 Z_j), \quad \ell = 0,1,2.$$

Since the sampling scheme of selecting the subcohort for a classical case-cohort study is sampling without replacement (see e.g. Self and Prentice, 1988), which is not an independent sampling, proofs of asymptotic properties are not the same as that for independent Bernoulli sampling. We thus discuss the asymptotic properties of the estimating equations (2) and (3) separately.

In the next two sections, we show that the estimators obtained from both estimating functions (2) and (3) are asymptotically normal under specified regularity conditions. Proofs follow the approach of Tsiatis (1990) and details are deferred to the Appendix. We first show asymptotic linearity of these estimating functions, then show asymptotic normality based on the linear approximation. Consistency is discussed briefly at the end of this section.

Remark: In an effort to improve efficiency, some authors include failures outside the subcohort in constructing $\widetilde{Z}^B$ in function (3). For example, they might take $W_i = \Delta_i + (1-\Delta_i)R_i/\pi(Z_i^*)$. We do not adopt this type of weights here for two reasons:

(i) With such weights, martingale theory does not apply since $\Delta_i \equiv N_i(T^*)$ is not predictable and $Z_i - \widetilde{Z}^B(t,\beta_0)$ is not adapted to the filtration $\mathcal{F}_n(t,\beta_0)$. So a different method of proof is needed.

(ii) For the Cox model, the estimator obtained from the counterpart of (2) is close to fully efficient in cases of primary interest where the disease rate is low and the subcohort size is not too small. See for example Figure 1 and related discussion in Nan, Emond and Wellner (2004) for details. We expect a similar feature in the censored linear regression model.

For the cohort data, we make the same assumptions (O) and (A)-(F) below as Tsiatis (1990). Conditions with respect to case-cohort designs will be given later. Without loss of generality, we assume that $\beta_0 = 0$ and denote $\bar{Z}(u) = \bar{Z}(u,0)$.

Assumption (O): The follow-up is truncated at a fixed T^* (with the same transformation as T and C), which satisfies the condition that for some $\xi > 0$

$$P(X_i \geq T^* + \xi) \geq \psi > 0 \quad \text{for all } i.$$

Assumption (A): The density function $f(x)$ of e_i is bounded by K_1 for $x \leq T^* + \xi$.

Assumption (B): The density function $g_i(x)$ of C_i is bounded by K_2 for all i and $x \leq T^* + \xi$.

Assumption (C): There exists a function $\theta(u)$ with $\int_{-\infty}^{T^*} |\theta(u)| \, du < \infty$ such that

$$|\lambda(u + \eta) - \lambda(u) - \eta\lambda'(u)| \leq \eta^2\theta(u)$$

for $u \leq T^*$ and $|\eta| \leq \xi$;

Assumption (D): The covariate Z has finite support. Without loss of generality, we suppose that $|Z_i| \leq 1$ for $i = 1, \cdots, n$.

Assumption (E): There exists a continuous function $\mu(u, \beta)$ and a neighborhood, $\mathcal{B}_0$, of $\beta = 0$, such that

$$\sup_{\beta \in \mathcal{B}_0, u \leq T^* + \xi} \left\{ \|\bar{Z}(u, \beta) - \mu(u, \beta)\| \right\} \to 0$$

in probability.

Assumption (F): There exists a continuous function $A(u, \beta)$ such that

$$\sup_{\beta \in \mathcal{B}_0, u \leq T^* + \xi} \left\| n^{-1} \sum_{i=1}^{n} \{Z_i - \bar{Z}(u, \beta)\}^2 Y_i(u + \beta Z_i) - A(u, \beta) \right\| \to 0$$

in probability.

Note: Evidently $\mu(u, \beta) = E\{ZY(u+\beta Z)\}/E\{Y(u+\beta Z)\}$ is bounded by Assumption (O). Conditions (E) and (F) need not be assumptions since they are actually a consequence of Assumptions (O) and (D) as can be shown using empirical processes theory. See Lemma 7.3 in Ritov (1990), where on page 310 we also see that $\mu(u, \beta_0) = E[Z|X - \beta_0 Z \geq u] = E[Z|X - \beta_0 Z = u, \Delta = 1]$. We include them as "assumptions" here for ease of reference. The notation $|\cdot|$ and $\|\cdot\|$ is adopted from Andersen and Gill (1982).

Let $g(0) = \int_{-\infty}^{T^*} A(u, 0)\lambda'(u) \, du$ and $S_n^*(\beta) = S_n(0) + n\beta g(0)$ for cohort data. Tsiatis (1990) showed that, for any $C > 0$,

$$\sup_{|\beta| \leq Cn^{-1/2}} n^{-1/2} |S_n(\beta) - S_n^*(\beta)| \to 0$$

in probability. In other words, $S_n(\beta)$ and $S_n^*(\beta)$ are asymptotically equivalent, and $\hat{\beta}$, the zero-crossing of $S_n(\beta)$, is asymptotically equivalent to β^* where $S_n^*(\beta^*) = 0$; that is, $n^{1/2}(\hat{\beta} -$

$\beta^*) \to 0$ in probability. Thus $\hat{\beta}$ has the same asymptotic distribution as β^*, which is normal by the martingale central limit theorem.

Define $\widetilde{S}_n^*(\beta) \equiv \widetilde{S}_n(0) + n\beta g(0)$ for the classical case-cohort design, and $\widetilde{S}_n^{B*}(\beta) \equiv \widetilde{S}_n^B(0) + n\beta g(0)$ for the stratified case-cohort design. Similar to the approach for cohort data described above, we will show that

$$\sup_{|\beta| \leq Cn^{-1/2}} n^{-1/2} |\widetilde{S}_n(\beta) - \widetilde{S}_n^*(\beta)| \to 0 \quad \text{and} \quad \sup_{|\beta| \leq Cn^{-1/2}} n^{-1/2} |\widetilde{S}_n^B(\beta) - \widetilde{S}_n^{B*}(\beta)| \to 0$$

in probability for any bounded value $C > 0$. Then, with some abuse of notation, the asymptotic distribution of $\hat{\beta}$, the zero crossing of $\widetilde{S}_n(\beta)$ or $\widetilde{S}_n^B(\beta)$, is the same as that of β^* where $\widetilde{S}_n^*(\beta^*) = 0$ or $\widetilde{S}_n^{B*}(\beta^*) = 0$. The asymptotic distribution of β^* is determined by asymptotic properties of $\widetilde{S}_n(0)$ or $\widetilde{S}_n^B(0)$. Proofs of the asymptotic normality of $n^{-1/2}\widetilde{S}_n(0)$ and $n^{-1/2}\widetilde{S}_n^B(0)$ are not as straightforward as for cohort data, but can be developed along the lines of Self and Prentice (1988).

The above arguments require that $\hat{\beta}$ is $n^{1/2}$-consistent. We provide a brief argument here without going into details. For our model settings and assumptions, the three estimating functions $n^{-1}S_n(\beta)$ (assuming $\omega_n(u, \beta) = 1$ for simplicity), $n^{-1}\widetilde{S}_n(\beta)$, and $n^{-1}\widetilde{S}_n^B(\beta)$ in (1), (2), and (3) respectively, converge to the same limit in probability. Hence their roots converge to the same point given that the limit is a continuous function of θ, which implies the estimators from case-cohort studies are consistent if the estimator from the cohort study is consistent. Furthermore, the difference between $n^{-1}S_n(\beta)$ and $n^{-1}\widetilde{S}_n^B(\beta)$ (or $n^{-1}\widetilde{S}_n(\beta)$) vanishes with $n^{1/2}$-rate, which can be argued using empirical process theory and, in particular, the Donsker property. Thus the estimators from case-cohort studies are $n^{1/2}$-consistent if the estimator from the cohort study is $n^{1/2}$-consistent.

3 The Classical Case-Cohort Design

Along with the assumptions in the previous section, the following additional conditions ensure the desired asymptotic properties for a classical case-cohort design:

Assumption (G): There exists a constant $\alpha \in (0, 1)$ such that $\tilde{n}/n \to \alpha$ in probability.

Assumption (H) Asymptotic stability of subcohort averages:

$$\frac{1}{\tilde{n}} \sum_{i \in \mathcal{C}} I(X_i \geq T^* + \xi) \to P(X_1 \geq T^* + \xi) \quad \text{in probability;} \tag{4}$$

$$\sup_{\beta \in \mathcal{B}_0, u \leq T^* + \xi} |\widetilde{D}^{(k)}(u, \beta) - d^{(k)}(u, \beta)| \to 0 \quad \text{in probability,} \quad k \in \{0, 1, 2\}; \tag{5}$$

and the sequence of distributions of $n^{1/2}\{\widetilde{Z}(u, 0) - \bar{Z}(u, 0)\}$ is tight on the product space of left continuous functions with right-hand limits equipped with the product Skorohod topology.

From (5) we have

$$\sup_{\beta \in \mathcal{B}_0, u \leq T^* + \xi} \left\{ \|\widetilde{Z}(u, \beta) - \mu(u, \beta)\| \right\} \to 0 \tag{6}$$

in probability. The above assumptions are part of Condition G in Self and Prentice (1988). We only have part of their G(ii), the tightness of $n^{1/2}\{\widetilde{Z}(u, 0) - \bar{Z}(u, 0)\}$, because our model automatically satisfies the rest of G(ii).

Following Tsiatis (1990), we first show point-wise convergence. That is, for any fixed d, we show that

$$n^{-1/2} |\widetilde{S}_n(n^{-1/2} d) - \widetilde{S}_n^*(n^{-1/2} d)| \to 0 \tag{7}$$

in probability. It follows that for a mesh $d_0, \cdots, d_m$ from $-C$ to C,

$$\max_{i \leq m} n^{-1/2} |\widetilde{S}_n(n^{-1/2} d_i) - \widetilde{S}_n^*(n^{-1/2} d_i)| \to 0$$

in probability. To complete the proof of uniform convergence, we show that $n^{-1/2}\widetilde{S}_n(\beta)$ as a function of β cannot fluctuate too much within any interval in the mesh. That is, for any $\epsilon > 0$, there exists a mesh size $\delta > 0$ such that

$$\lim_{n \to \infty} P \left\{ \sup_{dn^{-1/2} \leq \beta \leq (d+\delta) n^{-1/2}} n^{-1/2} |\widetilde{S}_n(\beta) - \widetilde{S}_n(dn^{-1/2})| \geq \epsilon \right\} = 0 \tag{8}$$

for any $|d| \leq C$.

Write

$$\widetilde{S}_n(\beta) = \sum_{i=1}^{n} \int_{-\infty}^{T^*} \{Z_i - \widetilde{Z}(u, \beta)\} \, dM_i(u + \beta Z_i) \tag{9}$$

$$+ \sum_{i=1}^{n} \int_{-\infty}^{T^*} Y_i(u + \beta Z_i) \{Z_i - \widetilde{Z}(u, \beta)\} \lambda(u + \beta Z_i) \, du \, . \tag{10}$$

For the point-wise convergence, we show that, for a fixed sequence of constants β_n converging to 0, (9) and (10) are asymptotically equivalent to $\widetilde{S}_n(0)$ and $\beta_n g(0)$ respectively. The following three lemmas are essentially Lemmas 3.1-3.3 in Tsiatis (1990) with $\bar{Z}$ replaced by $\widetilde{Z}$. The proofs need some extra effort for a case-cohort design and are given in the Appendix.

Lemma 1. Let β_n denote a fixed sequence of constants converging to 0. Then

$$n^{-1/2}\left[\sum_{i=1}^{n}\int_{-\infty}^{T^*}\{Z_i-\widetilde{Z}(u,\beta_n)\}\ dM_i(u+\beta_n Z_i)-\sum_{i=1}^{n}\int_{-\infty}^{T^*}\{Z_i-\mu(u,\beta_n)\}\ dM_i(u+\beta_n Z_i)\right] \quad (11)$$

converges to 0 in probability.

Lemma 2. Let β_n denote a fixed sequence of constants converging to 0. Then

$$n^{-1/2}\left[\sum_{i=1}^{n}\int_{-\infty}^{T^*}\{Z_i-\widetilde{Z}(u,\beta_n)\}\ dM_i(u+\beta_n Z_i)-\widetilde{S}_n(0)\right] \quad (12)$$

converges to 0 in probability.

Lemma 3. The integral (10) satisfies

$$n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{T^*}Y_i(u+\beta_n Z_i)\{Z_i-\widetilde{Z}(u,\beta_n)\}\lambda(u+\beta_n Z_i)\ du=\beta_n\{g(0)+o_p(1)\}$$

where $g(0)=\int_{-\infty}^{T^*}A(u,0)\lambda'(u)\ du$.

We thus have the following theorem:

Theorem 1. (Point-wise convergence.) The convergence in (7) is true.

Proof. With β and β_n replaced by $n^{-1/2}d$, by Lemma 2 we have $n^{-1/2}\{(9)-\widetilde{S}_n(0)\}\to 0$ in probability; and by Lemma 3 we have $n^{-1/2}\{(10)-n^{-1/2}dg(0)\}\to 0$ in probability. □

Theorem 2. (Uniform convergence.) The convergence in (8) is true.

The proof of Theorem 2 needs extra consideration for a case-cohort design. We give a proof in the Appendix.

Let $\hat{\beta}$ be the value of β where $\widetilde{S}_n(\beta)$ changes sign, which is in a $n^{1/2}$-neighborhood of 0. Let β^* be the solution to $\widetilde{S}_n^*(\beta)=0$, which is

$$n^{1/2}\beta^*=-n^{-1/2}\widetilde{S}_n(0)/g(0).$$

Then by Theorem 2 (using the similar arguments to Tsiatis (1990)) we have $n^{1/2}(\hat{\beta}-\beta^*) \to 0$ in probability. Thus, $n^{1/2}\hat{\beta}$ converges to the same limiting distribution as $n^{1/2}\beta^*$. Finally we show

Theorem 3. (Asymptotic normality.) The random variable $n^{-1/2}\widetilde{S}_n(0)$ is asymptotically normal with zero mean and, if $|g(0)| > 0$, $n^{1/2}\hat{\beta}$ is asymptotically normal with zero mean.

The proof of Theorem 3 follows Self and Prentice (1988), pages 72-73. Details are given in the Appendix.

4 Stratified Case-cohort Design

As outlined in Section 2, the subcohort $\mathcal{C}$ is selected by independent Bernoulli sampling where $R_i = I(i \in \mathcal{C})$ and $P(R_i = 1) = \pi(Z_i^*),\ i = 1, ..., n$. In this, Z_i^* is available for everyone but not involved in the regression model.

Analogous to the last section, we make the following assumption:

Assumption (G'): $\pi(Z_i^*) \geq \rho,\ i = 1, \cdots, n$ for some constant $\rho > 0$.

In the stratified case cohort design, the observed cohort data are independent and identically distributed and the results analogous to Assumption (H) follow from empirical processes theory. Thus we obtain the following stability conditions:

$$\|\widetilde{D}_B^{(k)}(u,\beta) \to d_B^{(k)}(u,\beta)\| \to 0 \quad \text{in probability for } k \in \{0,1,2\}\ , \tag{13}$$

and the result that the process $n^{1/2}\{\widetilde{Z}(u,0) - \bar{Z}(u,0)\}$ converges weakly to a zero mean Gaussian process with continuous sample path. Thus it is easily seen that Lemmas 1, 2, and 3 for $\widetilde{S}_n^B(\beta)$ hold and analogous to Theorem 1, we have

Theorem 4. (Point-wise convergence) For any fixed d,

$$n^{-1/2}|\widetilde{S}_n^B(n^{-1/2}d) - \widetilde{S}_n^{B*}(n^{-1/2}d)| \to 0$$

in probability.

The uniform convergence in the following Theorem 5 can be proved using the same idea as in previous section, which shows the asymptotic linearity of $\widetilde{S}_n^B(\beta)$. Details are given in the Appendix.

Theorem 5. (Uniform convergence.) For any $C > 0$,

$$\sup_{|\beta| \le Cn^{-1/2}} n^{-1/2} |\widetilde{S}_n^B(\beta) - \widetilde{S}_n^{B*}(\beta)| \to 0 \quad \text{in probability.} \tag{14}$$

Let $\hat{\beta}$ be the value of β where $\widetilde{S}_n^B(\beta)$ changes sign. As in the preceding section, the asymptotic distribution of $\hat{\beta}$ is the same as that of $-n^{-1/2}\widetilde{S}_n^B(0)/g(0)$, and the following theorem is proved in the appendix.

Theorem 6. (Asymptotic normality.) The random variable $n^{-1/2}\widetilde{S}_n^B(0)$ is asymptotically normal with zero mean and, if $|g(0)| > 0$, $n^{1/2}\hat{\beta}$ is asymptotically normal with zero mean.

5 The Variance Estimator

We use the method of Huang (2002) to calculate a variance estimator of $\hat{\beta}$. See also Kalbfleisch and Prentice (2002, page 238). Let $\Psi_n(\beta)$ denote either $\widetilde{S}_n(\beta)$ or $\widetilde{S}_n^B(\beta)$. Let $\Sigma(\beta_0)$ be the variance matrix for $n^{-1/2}\Psi_n(\beta_0)$. Suppose that $\Sigma(\hat{\beta}) = \mathbf{C}\mathbf{C}^T$ where $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_d)$. Let $\widetilde{\beta}_j$ satisfy $n^{-1/2}\Psi_n(\widetilde{\beta}_j) = \mathbf{c}_j$, $j = 1, \dots, d$. Let $\mathbf{D} = (\widetilde{\beta}_1 - \hat{\beta}, \dots, \widetilde{\beta}_d - \hat{\beta})$. Then $n\mathbf{D}\mathbf{D}^T$ is a consistent variance estimator of $n^{1/2}(\hat{\beta} - \beta_0)$.

We continue with calculations to find a variance estimator for $n^{-1/2}\Psi_n(\beta_0) = \widetilde{S}_n^B(\beta_0)$ in the stratified case and include some brief remarks about the classical case at the end of this section. We again assume $\beta_0 = 0$ without loss of generality. Equation (3) can be rewritten as

$$\begin{aligned} \widetilde{S}_n^B(0) &= n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{T^*} \{Z_i - \mu(u)\} dN_i(u) \\ &\quad - n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{T^*} \{\widetilde{Z}^B(u) - \mu(u)\} dN_i(u) \ . \end{aligned}$$

From $\sup_{u \le T^*} \|\widetilde{Z}^B(u) - \mu(u)\| = o_p(1)$ and martingale theory, the second term on the right hand side of the above equation is equal to

$$n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{T^*} \{\widetilde{Z}^B(u) - \mu(u)\} Y_i(u) d\Lambda(u) \tag{15}$$

plus a term $o_p(1)$ for underlying cohort data.

We now show that (15) can be replaced by

$$n^{-1/2} \sum_{i=1}^n \int_{-\infty}^{T^*} W_i \{\widetilde{Z}^B(u) - \mu(u)\} Y_i(u) d\Lambda(u), \tag{16}$$

which is fully determined by observed case-cohort data. Taking the difference of (15) and (16), we obtain

$$\begin{aligned}
|D_n| &= \left| \int_{-\infty}^{T^*} \left\{ n^{-1/2} \sum_{i=1}^{n} (1 - W_i) Y_i(u) \right\} \{ \widetilde{Z}^B(u) - \mu(u) \} d\Lambda(u) \right| \\
&\leq \int_{-\infty}^{T^*} \left| n^{-1/2} \sum_{i=1}^{n} (1 - W_i) Y_i(u) \right| \cdot \left| \widetilde{Z}^B(u) - \mu(u) \right| d\Lambda(u) \\
&\leq \sup_{u \leq T^*} \left| n^{-1/2} \sum_{i=1}^{n} (1 - W_i) Y_i(u) \right| \cdot \sup_{u \leq T^*} \left| \widetilde{Z}^B(u) - \mu(u) \right| \cdot \Lambda(T^*) \ .
\end{aligned}$$

Now, $h(u) \equiv (1-W)Y(u)$ with $|h(u)|$ bounded by $1+1/\rho$ forms a VC-class $\mathcal{H} = \{h(u) : u \leq T^*\}$ of index 2. Since $\mathcal{H}$ is a class of zero mean random variables, Theorem 2.6.7 and Theorem 2.14.9 in van der Vaart and Wellner (1996) imply that $\sup_{u \leq T^*} \left| n^{-1/2} \sum_{i=1}^{n} (1 - W_i) Y_i(u) \right|$ has exponential tail probability, and thus is equal to $O_p(1)$. Further, since $\sup_{u \leq T^*} \left| \widetilde{Z}^B(u) - \mu(u) \right| = o_p(1)$ and $\Lambda(T^*)$ is bounded, we have $|D_n| = o_p(1)$.

We rewrite (16) as

$$n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{T^*} W_i \{Z_i - \mu(u)\} Y_i(u) d\Lambda(u) \ ,$$

and find

$$\begin{aligned}
\widetilde{S}_n^B(0) &= n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{T^*} \{Z_i - \mu(u)\} dN_i(u) \\
&\quad - n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{T^*} W_i \{Z_i - \mu(u)\} Y_i(u) d\Lambda(u) + o_p(1) \ . \qquad (17)
\end{aligned}$$

It can be verified that $E\left[\int_{-\infty}^{T^*} \{Z - \mu(u)\} dN(u)\right] = E\left[\int_{-\infty}^{T^*} \{Z - \mu(u)\} Y(u) d\Lambda(u)\right] = 0$.

The right hand side of equation (17) is $n^{-1/2}$ times the sum of independent and identically distributed zero mean random variables with finite variance. Thus the asymptotic variance of $\widetilde{S}_n^B(0)$ can be estimated as the sample variance of the difference of the two integrals in (17) with $\mu(u)$ replaced by $\widetilde{Z}^B(u, \hat{\beta})$ and $\Lambda(u)$ by the Neslon-Aalen estimator $\hat{\Lambda}(u, \hat{\beta})$ as in Tsiatis (1990). This also provides an alternative proof of the asymptotic normality of $n^{-1/2}\widetilde{S}_n^B(0)$.

For a classical case-cohort design, we can follow the approach of Chen and Lo (1999, page 763) to estimate the asymptotic variance of $n^{-1/2}\widetilde{S}_n(0)$ with $\Sigma_1 + (n/\widetilde{n} - 1)\Sigma_2$, where Σ_1 is the sample variance of $\int_{-\infty}^{T^*} \{Z - \mu(u)\} dN(u)$ obtained using the cohort data and Σ_2 is the sample variance of $\int_{-\infty}^{T^*} \{Z - \mu(u)\} Y(u) d\Lambda(u)$ obtained using the subcohort only with $\mu(u)$ replaced by $\widetilde{Z}(u, \hat{\beta})$ and $\Lambda(u)$ by $\hat{\Lambda}(u, \hat{\beta})$.

6 Numerical Examples

As is well known, root finding of a non-smooth function can be challenging. In the univariate case, the bisection algorithm or the false position method can be used (Press, Teukolsky, Vetterling, and Flannery (2002), Chapter 9). Neither method requires derivatives of the function. By connecting two points where the function has different signs, the false position method takes the next approximation to the root as the point where the resulting line crosses the axis. In most instances, the false position method converges faster than bisection.

The multi-dimensional problem can be reduced to the one-dimensional case by recursively applying the univariate method (e.g., see Huang, 2002). Thus, we iteratively find root for β_1 through β_d using the false position method while holding other β's constant. The algorithm stops when the change in each component of β is smaller than a pre-specified quantity (e.g., 10^{-5}). The method seems to work well in our numerical examples.

We report on simulations to evaluate the finite sample performance of the proposed estimators. For these, we took $Z \sim \text{Bernoulli}(p_Z)$, $\epsilon \sim N(0,1)$, failure time T from the model $T = \beta Z + \epsilon$, and censoring time C where $\exp(C) \sim \text{Exponential}(\lambda)$ where λ is chosen so that the censoring rate is about 90%. We chose $p_Z = 0.1$ or $p_Z = 0.3$, and $\beta = 0$ or $\beta = 1$. In addition, we defined the distribution of Z^* using $\eta = P(Z^* = 1|Z = 1)$ and $\nu = P(Z^* = 0|Z = 0)$. We chose (η, ν) : $(\eta, \nu) \in \{(0.5, 0.5), (0.7, 0.7), (0.9, 0.9)\}$. Thus $Z^* \sim \text{Bernoulli}((1-\nu)(1-p_Z) + \eta p_Z)$. The subcohort is either a simple random sample of the cohort or a stratified sample selected by independent Bernoulli sampling with selection probability $\pi(Z^*)$ chosen so that approximately equal numbers of subjects are selected from the two strata, $\{Z^* = 1\}$ and $\{Z^* = 0\}$. In all cases, the size of the subcohort is 10% of the entire cohort. Simulation results that compare the classical, stratified and full cohort are given in Table 1.

In the cases considered, the biases of all the methods are minimal. The classical case-cohort design does slightly better than that with subcohort selected by independent Bernoulli sampling when Z and Z^* are uncorrelated ($\eta = \nu = 0.5$). As the correlation of Z and Z^* increases, the efficiency of the stratified case-cohort design increases.

We also analyze a data set collected in two randomized clinical trials in Wilms tumor patients to illustrate the application of the censored linear regression model for case-cohort studies. These two trials were the third and the fourth trials conducted by the National

Wilms Tumor Study Group (NWTSG) (see e.g. D'Angio, Breslow, Beckwith et al. (1989); Green, Breslow, Beckwith et al. (1998)). We considered linear models for the logarithm of the time to tumor relapse with covariates tumor histologic type (favorable versus unfavorable), tumor stage (I-IV), trial (NWTSG-3 versus NWTSG-4), the logarithm of tumor diameter, and age at cancer diagnosis. Tumor histology was assessed in two ways: a 'local histology' determined at each individual site and 'central histology' evaluated at a central facility. The central histology is regarded as the gold standard, but the reevaluation process was expensive and time consuming. The proportion of unfavorable central histology was 11%. The sensitivity and specificity of unfavorable local histology was 78% and 93%, respectively. Although this example is done solely for illustration, one advantage of the subcohort analyses is that the central histology would only need to be done for cases who fail or are in the subcohort.

There were 4222 patients in the data set of whom 4117 had sufficient data to include in the analysis. Local histology on three levels (favorable, unfavorable and unknown) was used to define the stratified subcohort design; there were 216 patients with unknown local histologic type. Our analyses estimated the effect of the central histologic type (adjusted for other covariates) using data from the original cohort, a classical case-cohort design, and a stratified case-cohort design. In the whole cohort, 727 patients are observed to relapse. The subcohort size is targeted at 800 patients; in the stratified analysis, we selected respectively about 360, 360 and 80 from the favorable, unfavorable and unknown local histology groups. For the case cohort analyses, we re-sampled the subcohort 20 times and reported mean estimated parameters and mean standard error estimators in Table 2.

The first row is the result from the full-cohort data analysis and all factors are highly significant except the log tumor diameter. All the mean estimated parameters from the case-cohort studies are close to the full-cohort estimates. The stratified analysis has slightly smaller standard errors for central unfavorable histology than the classical case-cohort analysis, but it has little effect on the standard errors of other estimates. This finding is very similar to that in Kulich and Lin (2000), where an additive hazard model is used.

7 Discussion

Due to the fact that estimating functions for β in censored linear regressions are step functions, developing efficient and reliable computational methods for $\hat{\beta}$ has been a challenging problem. Lin, Wei, and Ying (1998) noted that for cohort data, the estimating function $S_n(\omega_n, \beta)$ in (1) with Gehan weights is the gradient of a objective function that can be minimized by linear programming. Jin, Lin, Wei, and Ying (2003) extended their method to estimating functions with arbitrary weights using Gehan estimator as the initial value. It would be interesting to know whether linear programming is applicable to case-cohort studies.

As mentioned in the remark under equation (3), it is tempting to include failures outside the subcohort into the estimating function for a case-cohort study to improve efficiency. Apparently martingale theory fails in this case and it would be worthwhile to develop rigorous proofs for corresponding asymptotic properties.

Acknowledgement

We thank the National Wilms Tumor Study Group for providing the data.

Appendix

Proof of Lemma 1. Note that (11) is equal to $R(T^*)$ where

$$R(u) = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{u} \{\widetilde{Z}(u, \beta_n) - \mu(u, \beta_n)\} \, dM_i(u + \beta_n Z_i)$$

is an $\mathcal{F}_n(u, \beta_n) \vee \sigma(\mathcal{C})$ martingale and $\sigma(\mathcal{C})$ denotes the σ-algebra of possible subcohort selections (see Self and Prentice, 1988). The proof then follows the proof of Lemma 3.1 in Tsiatis (1990) with $\bar{Z}$ replaced by $\widetilde{Z}$ and applying condition (6).

Proof of Lemma 2. Write (12) as

$$\begin{aligned}
n^{-1/2} \Bigg[\sum_{i=1}^n \int_{-\infty}^{T^*} \{Z_i - \widetilde{Z}(u,\beta_n)\}\ dM_i(u+\beta_n Z_i) \\
- \sum_{i=1}^n \int_{-\infty}^{T^*} \{Z_i - \mu(u,\beta_n)\}\ dM_i(u+\beta_n Z_i) \Bigg] \\
+ n^{-1/2} \Bigg[\sum_{i=1}^n \int_{-\infty}^{T^*} \{Z_i - \mu(u,\beta_n)\}\ dM_i(u+\beta_n Z_i) \\
- \sum_{i=1}^n \int_{-\infty}^{T^*} \{Z_i - \mu(u,0)\}\ dM_i(u) \Bigg] \qquad (18) \\
+ n^{-1/2} \Bigg[\sum_{i=1}^n \int_{-\infty}^{T^*} \{Z_i - \mu(u,0)\}\ dM_i(u) - \widetilde{S}_n(0) \Bigg].
\end{aligned}$$

By Lemma 1, we have the first and third terms converge in probability to 0.

The term (18) is exactly the same as (3.9) in Lemma 3.2 of Tsiatis (1990). Readers can refer there for the rest of the proof. $\square$.

Proof of Lemma 3. We first show that

$$\begin{aligned}
n^{-1} \sum_{i=1}^n \int_{-\infty}^{T^*} Y_i(u+\beta_n Z_i)\{Z_i - \widetilde{Z}(u,\beta_n)\}\lambda(u+\beta_n Z_i)\ du \\
= n^{-1} \sum_{i=1}^n \int_{-\infty}^{T^*} Y_i(u+\beta_n Z_i)\{Z_i - \bar{Z}(u,\beta_n)\}\lambda(u+\beta_n Z_i)\ du + o_p(1). \qquad (19)
\end{aligned}$$

Lemma 3.3 in Tsiatis (1990) completes the proof.

To show (19), we show

$$n^{-1} \sum_{i=1}^n \int_{-\infty}^{T^*} Y_i(u+\beta_n Z_i)\{\widetilde{Z}(u,\beta_n) - \bar{Z}(u,\beta_n)\}\lambda(u+\beta_n Z_i)\ du = o_p(1).$$

By Assumption (E) and condition (6), there exists $n(\epsilon, K)$ such that for any $n > n(\epsilon, K)$,

$$\begin{aligned}
& P\Big\{ \sup_{u\le T^*} |\widetilde{Z}(u,\beta_n) - \bar{Z}(u,\beta_n)| > K \Big\} \\
\le\ & P\Big\{ \sup_{u\le T^*} |\widetilde{Z}(u,\beta_n) - \mu(u,\beta_n)| + \sup_{u\le T^*} |\bar{Z}(u,\beta_n) - \mu(u,\beta_n| > K \Big\} \\
\le\ & P\Big\{ \sup_{u\le T^*} |\widetilde{Z}(u,\beta_n) - \mu(u,\beta_n)| > K/2 \Big\} + P\Big\{ \sup_{u\le T^*} |\bar{Z}(u,\beta_n) - \mu(u,\beta_n| > K/2 \Big\} \\
<\ & \epsilon.
\end{aligned}$$

Since $|\beta_n Z_i| \leq |\beta_n| \rightarrow 0$, the integral above cannot exceed $K\Lambda(T^* + \xi)$ as long as n is large enough that $|\beta_n|$ is less than ξ. By Assumption (A), we have $\Lambda(T^* + \xi) = -\log\{S(T^* + \xi)\}$ finite. Hence if we choose $K \leq \delta/\Lambda(T^* + \xi)$, then

$$P\left[n^{-1}\sum_{i=1}^{n}\int_{-\infty}^{T^*} Y_i(u + \beta_n Z_i)\{\widetilde{Z}(u, \beta_n) - \bar{Z}(u, \beta_n)\}\lambda(u + \beta_n Z_i)\ du > \delta\right] < \epsilon$$

for $n > n(\epsilon, K)$. $\square$

Proof of Theorem 2. We put a probabilistic bound on the maximum change of the statistic $\widetilde{S}_n(\beta)$ as β varies from $dn^{-1/2}$ to $(d+\delta)n^{-1/2}$. Since

$$\widetilde{S}_n(\beta) = \sum_{i=1}^{n}\int_{-\infty}^{T^*}\{Z_i - \widetilde{Z}(u, \beta)\}\ dN_i(u + \beta Z_i) = \sum_{i\in\mathcal{C}\cup\mathcal{D}}\int_{-\infty}^{T^*}\{Z_i - \widetilde{Z}(u, \beta)\}\ dN_i(u + \beta Z_i)$$

with $N_i(u + \beta Z_i) = I\{X_i \leq u + \beta Z_i, \Delta_i = 1\}$ and

$$\widetilde{Z}(u, \beta) = \sum_{j\in\mathcal{C}} Z_j Y_j(u + \beta Z_j)\Big/\sum_{j\in\mathcal{C}} Y_j(u + \beta Z_j)\ .$$

We can write $\widetilde{S}_n(\beta)$ as

$$\sum_{i\in\mathcal{C}\cup\mathcal{D}}\Delta_i\left[Z_i - \frac{\sum_{j\in\mathcal{C}} Z_j I(X_j - \beta Z_j \geq X_i - \beta Z_i)}{\sum_{j\in\mathcal{C}} I(X_j - \beta Z_j \geq X_i - \beta Z_i)}\right]\ .$$

Hence the statistic depends only on ranks of residuals $X_i - \beta Z_i$, $i \in \mathcal{C}\cup\mathcal{D}$. As β varies from $dn^{-1/2}$ to $(d+\delta)n^{-1/2}$, the statistic changes whenever there is a change in the ranks of residuals $X_i - \beta Z_i$ for $i \in \mathcal{C}\cup\mathcal{D}$.

The maximum change can be bounded by the product of total number of pairs of ranks that will be interchanged and the maximum change of the statistic for each such interchange. First consider the maximum change of the statistic at an interchange. When β increases, the interchange in ranks occurs only between neighboring order statistics of the residuals $X_i - \beta Z_i$, $i \in \mathcal{C}\cup\mathcal{D}$. Let (i) denote the ith order statistic of residuals $X_i - \beta Z_i$, $i \in \mathcal{C}\cup\mathcal{D}$, and let $\widetilde{\mathcal{R}}\{(i), \beta\}$ be the risk set corresponding to (i) in the subcohort $\widetilde{C}$. Then $\widetilde{S}_n(\beta)$ can be written as

$$\sum_{i\in\mathcal{C}\cup\mathcal{D}}\Delta_{(i)}\left[Z_{(i)} - \widetilde{Z}_{(i)}(u, \beta)\right] \tag{20}$$

where

$$\widetilde{Z}_{(i)}(u, \beta) = \frac{\sum_{j\in\widetilde{\mathcal{R}}\{(i),\beta\}} Z_j}{N_{\widetilde{\mathcal{R}}\{(i),\beta\}}}\ .$$

Here $N_{\widetilde{\mathcal{R}}\{(i),\beta\}}$ is the number of subjects in the risk set $\widetilde{\mathcal{R}}\{(i),\beta\}$.

Now suppose as β increases to β^+, a change of order between (i) and $(i+1)$ occurs. If both (i) and $(i+1)$ are censored, then $\widetilde{S}_n(\beta)$ does not change. We need only consider the corresponding change in (20) for the following three cases:

1. Subject (i) is an event and $(i+1)$ is censored. Then $(i+1) \in \mathcal{C}$. The corresponding $\widetilde{Z}$ after interchange is $\widetilde{Z}_{(i)}(u,\beta^+)$ and

$$
\begin{aligned}
|\widetilde{S}_n(\beta) - \widetilde{S}_n(\beta^+)| &= |\widetilde{Z}_{(i)}(u,\beta^+) - \widetilde{Z}_{(i)}(u,\beta)| \\
&= \left| \frac{\sum_{j\in\widetilde{\mathcal{R}}\{(i),\beta\}} Z_j - Z_{(i+1)}}{N_{\widetilde{\mathcal{R}}\{(i),\beta\}} - 1} - \frac{\sum_{j\in\widetilde{\mathcal{R}}\{(i),\beta\}} Z_j}{N_{\widetilde{\mathcal{R}}\{(i),\beta\}}} \right| \\
&= \left| \frac{\sum_{j\in\widetilde{\mathcal{R}}\{(i),\beta\}} Z_j - N_{\widetilde{\mathcal{R}}\{(i),\beta\}} Z_{(i+1)}}{N_{\widetilde{\mathcal{R}}\{(i),\beta\}}\left[N_{\widetilde{\mathcal{R}}\{(i),\beta\}} - 1\right]} \right| \\
&\le \frac{2}{N_{\widetilde{\mathcal{R}}\{(i),\beta\}} - 1} \\
&= \frac{2}{N_{\widetilde{\mathcal{R}}\{(i),\beta^+\}}} .
\end{aligned}
$$

2. Subject (i) is censored and $(i+1)$ is an event. Then $(i) \in \mathcal{C}$, and thus

$$
\widetilde{Z}_{(i+1)}(u,\beta^+) = \frac{\sum_{j\in\widetilde{\mathcal{R}}\{(i+1),\beta\}} Z_j + Z_{(i)}}{N_{\widetilde{\mathcal{R}}\{(i+1),\beta\}} + 1} .
$$

Similar calculation yields

$$
|\widetilde{S}_n(\beta) - \widetilde{S}_n(\beta^+)| = |\widetilde{Z}_{(i+1)}(u,\beta^+) - \widetilde{Z}_{(i+1)}(u,\beta)| \le \frac{2}{N_{\widetilde{\mathcal{R}}\{(i+1),\beta^+\}}} .
$$

3. Both Subjects (i) and $(i+1)$ are events. Then

$$
\widetilde{Z}_{(i)}(u,\beta^+) = \begin{cases} \widetilde{Z}_{(i)}(u,\beta) & \text{if } (i+1) \notin \mathcal{C} \\ \dfrac{\sum_{j\in\widetilde{\mathcal{R}}\{(i),\beta\}} Z_j - Z_{(i+1)}}{N_{\widetilde{\mathcal{R}}\{(i),\beta\}} - 1} & \text{if } (i+1) \in \mathcal{C} , \end{cases}
$$

$$
\widetilde{Z}_{(i+1)}(u,\beta^+) = \begin{cases} \widetilde{Z}_{(i+1)}(u,\beta) & \text{if } (i) \notin \mathcal{C} \\ \dfrac{\sum_{j\in\widetilde{\mathcal{R}}\{(i+1),\beta\}} Z_j + Z_{(i)}}{N_{\widetilde{\mathcal{R}}\{(i+1),\beta\}} + 1} & \text{if } (i) \in \mathcal{C} , \end{cases}
$$

and

$$|\widetilde{S}_n(\beta) - \widetilde{S}_n(\beta^+)| = |\widetilde{Z}_{(i)}(u, \beta^+) + \widetilde{Z}_{(i+1)}(u, \beta^+) - \widetilde{Z}_{(i)}(u, \beta) - \widetilde{Z}_{(i+1)}(u, \beta)| \ .$$

Straightforward calculation shows that

3a. if $(i) \notin \mathcal{C}$, $(i+1) \notin \mathcal{C}$, then $\widetilde{S}_n(\beta) - \widetilde{S}_n(\beta^+) = 0$;

3b. if $(i) \in \mathcal{C}$, $(i+1) \notin \mathcal{C}$, then

$$|\widetilde{S}_n(\beta) - \widetilde{S}_n(\beta^+)| \leq \frac{2}{N_{\widetilde{\mathcal{R}}\{(i+1),\beta^+\}}} \ ;$$

3c. if $(i) \notin \mathcal{C}$, $(i+1) \in \mathcal{C}$, then

$$|\widetilde{S}_n(\beta) - \widetilde{S}_n(\beta^+)| \leq \frac{2}{N_{\widetilde{\mathcal{R}}\{(i),\beta^+\}}} \ ;$$

3d. if $(i) \in \mathcal{C}$, $(i+1) \in \mathcal{C}$, then

$$|\widetilde{S}_n(\beta) - \widetilde{S}_n(\beta^+)| \leq \frac{2}{N_{\widetilde{\mathcal{R}}\{(i),\beta^+\}}} + \frac{2}{N_{\widetilde{\mathcal{R}}\{(i+1),\beta^+\}}}$$

Then the change in (20) for any single interchange of the ordered residuals is bounded by $4/N_{\widetilde{\mathcal{R}}\{(i+1),\beta\}}$. Since $P(X_i \geq T^* + \xi) \geq \psi > 0$ for the same ξ and all i, there exists n large enough that $\tilde{n}^{-1}\sum_{i\in\widetilde{C}} I(X_i \geq T^* + \beta_n Z_i) > \psi/2$ with probability close to 1. Then the number of residuals at risk will exceed $\tilde{n}\psi/2$ whenever an interchange in ranks takes place. Hence with arbitrarily large probability, the change is bounded by $8/(\psi\tilde{n})$.

In his Theorem 3.2, Tsiatis (1990) considered the total number of interchanges M for the underlying complete data that occur as β increases from $dn^{-1/2}$ to $(d+\delta)n^{-1/2}$. For any given $\epsilon > 0$, he showed that there exists $\delta > 0$ such that

$$\lim_{n\to\infty} P\{n^{-3/2}M \geq \epsilon\} = 0 \ .$$

Since the total number of order interchanges for case-cohort data is no greater than M, we have

$$\begin{aligned}\sup_{dn^{-1/2}\leq\beta\leq(d+\delta)n^{-1/2}} n^{-1/2}|\widetilde{S}_n(\beta) - \widetilde{S}_n(dn^{-1/2})| &\leq n^{-1/2}M \times \{8/(\psi\tilde{n})\} \\ &= n^{-3/2}M \times \{8/(\psi\alpha)\} + o(n^{-3/2})\end{aligned}$$

and we thus have proved the theorem. □

Proof of Theorem 3. From the definition of $\widetilde{S}_n(\beta)$, we have

$$\begin{aligned}
n^{-1/2}\widetilde{S}_n(0) &= n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{T^*}\{Z_i - \widetilde{Z}(u,0)\}\ dN_i(u) \\
&= n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{T^*}\{Z_i - \bar{Z}(u,0)\}\ dM_i(u) \\
&\qquad - n^{-1/2}\sum_{i=1}^{n}\int_{-\infty}^{T^*}\{\widetilde{Z}(u,0) - \bar{Z}(u,0)\}\ dM_i(u) \\
&\qquad - n^{-1/2}\int_{-\infty}^{T^*}\{\widetilde{Z}(u,0) - \bar{Z}(u,0)\}\ d\bar{\Lambda}(u)\ , \qquad (21)
\end{aligned}$$

where $\bar{\Lambda}(u) = \sum_{i=1}^{n}\Lambda(u)$. The second term in (21) converges to zero as in Self and Prentice (1988, pages 70-71), since it is a martingale with covariation process converging to zero. Then the proof of asymptotic normality of the sum of the first and third terms follows exactly as the proof of asymptotic normality of the score statistic (Theorem 3.1 in Self and Prentice, 1988). The following is the verification of condition (2) of their Proposition 1.

It is easily seen that

$$\begin{aligned}
&n^{1/2}\big\{\widetilde{Z}(t,0) - \bar{Z}(t,0)\big\} \\
&\quad = n^{1/2}\left\{\frac{\widetilde{D}^{(1)}(t,0)}{\widetilde{D}^{(0)}(t,0)} - \frac{D^{(1)}(t,0)}{D^{(0)}(t,0)}\right\} \\
&\quad = n^{1/2}\widetilde{D}^{(0)}(t,0)^{-1}\Big[\big\{\widetilde{D}^{(1)}(t,0) - D^{(1)}(t,0)\big\} - \big\{\widetilde{D}^{(0)}(t,0) - D^{(0)}(t,0)\big\}\bar{Z}(t,0)\Big] \\
&\quad = n^{1/2}d^{(0)}(t,0)^{-1}\Big[\big\{\widetilde{D}^{(1)}(t,0) - D^{(1)}(t,0)\big\} - \big\{\widetilde{D}^{(0)}(t,0) - D^{(0)}(t,0)\big\}\mu(t,0)\Big] + o_p(1)\ .
\end{aligned}$$

Let $f_{jn}(\mathbf{X}_n) = d^{(0)}(t,0)^{-1}[Z_jY_j(t) - \mu(t,0)Y_j(t)]$. Then the absolute value $|f_{jn}(\mathbf{X}_n)|$ is bounded by $d^{(0)}(t,0)^{-1}[1 + |\mu(t,0)|] \le 2d^{(0)}(t,0)^{-1}$ due to the fact that $|Z_i| \le 1$, $|Y_i(t)| \le 1$. With this choice of $f_{jn}(\mathbf{X}_n)$, $n^{1/2}\big\{\widetilde{Z}(t,0) - \bar{Z}(t,0)\big\}$ becomes $h_n(\mathbf{X}_n, \boldsymbol{\delta}_n)$ in Proposition 1 in Self and Prentice (1998). Here we adopt their notation $\mathbf{X_n}$ and $\boldsymbol{\delta}_n$ to denote cohort data and subcohort membership indicators, respectively. Since $d^{(0)}(t,0)$ is bounded away from 0, for any fixed ϵ, when $n^{1/2}\epsilon \ge 4d^{(0)}(t,0)^{-1}$, we have

$$n^{-1}\sum_{j=1}^{n}[f_{jn}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)]^2 I_{\{|f_{jn}(\mathbf{X}_n) - f_{\cdot n}(\mathbf{X}_n)| > n^{1/2}\epsilon\}} = 0\ ,$$

which verifies the condition (2) in the proposition.

By their Proposition 1 we know that the first and third terms in (21) are asymptotically independent. Thus $n^{-1/2}\widetilde{S}_n(0)$ converges to a normal distribution with variance coming from two parts, the asymptotic variance of $n^{-1/2}S_n(0)$ for underlying complete data and the variance from sampling. $\square$

Proof of Theorem 5. we again consider the maximum change of the statistic at an interchange and the number of interchanges. Write $\widetilde{S}_n^B(\beta)$ as

$$\sum_{i=1}^{n} \Delta_{(i)} \left[Z_{(i)} - \widetilde{Z}_{(i)}^B(u,\beta) \right] = \sum_{i \in \mathcal{C} \cup \mathcal{D}} \Delta_{(i)} \left[Z_{(i)} - \widetilde{Z}_{(i)}^B(u,\beta) \right] , \tag{22}$$

where

$$\widetilde{Z}_{(i)}^B(u,\beta) = \frac{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i),\beta\}} W_j Z_j}{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i),\beta\}} W_j} .$$

Here $\widetilde{\mathcal{R}}^B\{(i),\beta\}$ is the risk set for subjects in $\mathcal{C}$. We have $1 \le W_j \le 1/\rho$ for all $j \in \mathcal{C}$.

Now suppose as β increases to β^+, a change of order between (i) and $(i+1)$ occurs. We consider the corresponding change in (22) for all the completely observed subjects for the following three cases. The calculations are similar to that in the proof of Theorem 2.

1. Subject (i) is an event and $(i+1)$ is censored. Then $(i+1) \in \mathcal{C}$, and

$$\widetilde{Z}_{(i)}^B(u,\beta^+) = \frac{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i)\}} W_j Z_j - W_{(i+1)} Z_{(i+1)}}{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i)\}} W_j - W_{(i+1)}} .$$

Hence

$$|\widetilde{S}_n^B(\beta) - \widetilde{S}_n^B(\beta^+)| = |\widetilde{Z}_{(i)}^B(u,\beta^+) - \widetilde{Z}_{(i)}^B(u,\beta)| \le \frac{2W_{(i+1)}}{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i)\}} W_j - W_{(i+1)}} .$$

2. Subject (i) is censored and $(i+1)$ is an event. So $(i) \in \mathcal{C}$. We have

$$\widetilde{Z}_{(i+1)}^B(u,\beta^+) = \frac{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i+1)\}} W_j Z_j + W_{(i)} Z_{(i)}}{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i+1)\}} W_j + W_{(i)}} ,$$

and thus

$$|\widetilde{S}_n^B(\beta) - \widetilde{S}_n^B(\beta^+)| = |\widetilde{Z}_{(i+1)}^B(u,\beta^+) - \widetilde{Z}_{(i+1)}^B(u,\beta)| \le \frac{2W_{(i)}}{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i+1)\}} W_j + W_{(i)}} .$$

3. Both Subjects (i) and $(i+1)$ are events. Then

$$\widetilde{Z}_{(i)}^B(u,\beta^+) = \frac{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i)\}} W_j Z_j - W_{(i+1)} Z_{(i+1)}}{\sum_{j \in \widetilde{\mathcal{R}}^B\{(i)\}} W_j - W_{(i+1)}} ,$$

and

$$\widetilde{Z}^B_{(i+1)}(u,\beta^+) = \frac{\sum_{j\in\widetilde{\mathcal{R}}^B\{(i+1)\}} W_j Z_j + W_{(i)} Z_{(i)}}{\sum_{j\in\widetilde{\mathcal{R}}^B\{(i+1)\}} W_j + W_{(i)}} \ .$$

Hence

$$\begin{aligned}
|\widetilde{S}^B_n(\beta) - \widetilde{S}^B_n(\beta^+)| &= |\widetilde{Z}^B_{(i)}(u,\beta^+) + \widetilde{Z}^B_{(i+1)}(u,\beta^+) - \widetilde{Z}^B_{(i)}(u,\beta) - \widetilde{Z}^B_{(i+1)}(u,\beta)| \\
&\le |\widetilde{Z}^B_{(i)}(u,\beta^+) - \widetilde{Z}^B_{(i)}(u,\beta)| + |\widetilde{Z}^B_{(i+1)}(u,\beta^+) - \widetilde{Z}^B_{(i+1)}(u,\beta)| \\
&\le \frac{2W_{(i+1)}}{\sum_{j\in\widetilde{\mathcal{R}}^B\{(i)\}} W_j - W_{(i+1)}} + \frac{2W_{(i)}}{\sum_{j\in\widetilde{\mathcal{R}}^B\{(i+1)\}} W_j + W_{(i)}} \ .
\end{aligned}$$

We see that the change in (22) for any single interchange of the ordered residuals is bounded by

$$\frac{2(W_{(i)} + W_{(i+1)})}{\sum_{j\in\widetilde{\mathcal{R}}^B\{(i+1)\}} W_j} \ \le \ \frac{4}{\rho N_{\widetilde{\mathcal{R}}^B\{(i+1)\}}} \ .$$

By similar argument to the classical case-cohort design, we have that the change is bounded by $8/(\tilde{n}^*\psi\rho) \le 8/(\tilde{n}\psi\rho)$. The rest of the proof is the same as that for the classical case-cohort design. $\square$

Prior to the proof of Theorem 6, we establish the following Lemma 4 that is a modification of Proposition 1 in Self and Prentice (1988).

Lemma 4: Denote $\mathbf{X}_n = (X_{1n}, \ldots, X_{nn})$ and $\mathbf{W}_n = (W_{1n}, \ldots, W_{nn})$. Let pairs (X_{1n}, W_{1n}), $\ldots$, (X_{nn}, W_{nn}) be independent and identically distributed random variables such that:
(1) $E[W_{1n}|X_{1n}] = E[W_{1n}] = 1$.
(2) For some scalar functions of $\mathbf{X}_n$, $f_{in}(\mathbf{X}_n)$, and for any $\epsilon > 0$,

$$\frac{1}{n}\sum_{i=1}^{n} E\left[f_{in}(\mathbf{X}_n)^2 (W_{in}-1)^2 I\{|f_{in}(\mathbf{X}_n)(W_{in}-1)| > n^{1/2}\epsilon\} \Big| \mathbf{X_n} \right] \to 0$$

in probability and

$$S^2_{fn} \equiv \frac{1}{n}\sum_{i=1}^{n} f_{in}(\mathbf{X}_n)^2 \mathrm{Var}(W_{in}|X_{in}) \to \sigma^2_f > 0$$

in probability.
(3) The scalar functions of $\mathbf{X}_n$, $g_n(\mathbf{X}_n)$, converge in distribution to a Gaussian random variable with mean zero and variance σ^2_g.
Then for $h_n(\mathbf{X}_n, \mathbf{W}_n) = n^{-1/2}\sum_{i=1}^{n} f_{in}(\mathbf{X}_n)(W_{in}-1)$, we have

$$\left(\begin{array}{c} g_n(\mathbf{X}_n) \\ h_n(\mathbf{X}_n, \mathbf{W}_n) \end{array} \right) \to N(0,\Sigma) \quad \text{in distribution with } \Sigma = \left[\begin{array}{cc} \sigma^2_g & 0 \\ 0 & \sigma^2_f \end{array} \right].$$

Proof. Let (g_n, h_n) denote $(g_n(\mathbf{X}_n), h_n(\mathbf{X}_n, \mathbf{W}_n))$. Let $h_n^* = h_n/S_{fn}$ and $v_n^* = v/S_{fn}$. Then

$$\begin{aligned} P(g_n \leq w, h_n \leq v) &= P(g_n \leq w, h_n^* \leq v^*) \\ &= E[I(g_n \leq w)P(h_n^* \leq v_n^* | \mathbf{X}_n)]. \end{aligned}$$

By conditions (1) and (2), Linderberg-Feller central limit theorem applies and thus we have

$$|P(h_n^* \leq v_n^* | \mathbf{X}_n) - \Phi(v^*)| \rightarrow_p 0,$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal random variable. Since $v^* \rightarrow_p v/\sigma_f$ and Φ is continuous, we have $\Phi(v^*) \rightarrow_p \Phi(v/\sigma_f)$. Thus by applying Dominated Convergence Theorem we have

$$\begin{aligned} P(g_n \leq w, h_n \leq v) &= E[I(g_n \leq w)\{\Phi(v/\sigma_f) + o_p(1)\}] \\ &= P(g_n \leq w)\Phi(v/\sigma_f) + o(1) \\ &\rightarrow \Phi(w/\sigma_g)\Phi(v/\sigma_f) \ . \end{aligned}$$

□

Proof of Theorem 6. As in equation (21), we have the following decomposition:

$$\begin{aligned} n^{-1/2}\widetilde{S}_n^B(0) &= n^{-1/2}\sum_{i=1}^n \int_{-\infty}^{T^*} \{Z_i - \widetilde{Z}^B(u,0)\} \, dN_i(u) \\ &= n^{-1/2}\sum_{i=1}^n \int_{-\infty}^{T^*} \{Z_i - \bar{Z}(u,0)\} \, dM_i(u) \\ &\qquad - n^{-1/2}\sum_{i=1}^n \int_{-\infty}^{T^*} \{\widetilde{Z}^B(u,0) - \bar{Z}(u,0)\} \, dM_i(u) \\ &\qquad - n^{-1/2}\int_{-\infty}^{T^*} \{\widetilde{Z}^B(u,0) - \bar{Z}(u,0)\} \, d\bar{\Lambda}(u) \ , \end{aligned}$$

Similar to the proof for the classical case-cohort design, the middle term in the above decomposition converges to zero. We can further show that the first and third terms converge jointly to independent Gaussian random variables by Lemma 4. Details are very close to the proof of Theorem 3 for non-trivial cases with $P\{\pi(Z^*) < 1\} > 0$ and thus omitted. □

REFERENCES

Andersen, P. K. and Gill, R. D. (1982), Cox's regression model for counting processes: a large sample study. *Annals of Statistics*, 10, 1100-1120.

Chen, K. and Lo, S.-H. (1999), Case-cohort and case-control analysis with Cox's model. *Biometrika*, 86, 755-764.

D'Angio, G. J., Breslow, N. E., Beckwith, J. B. et al. (1989), Treatment of Wilms tumor: results of the third National Wilms Tumor Study. *Cancer*, 64, 349-360.

Green, D. M., Breslow, N. E., Beckwith, J. B. et al. (1998), Comparison between single-dose and divided-dose administration of dactinomycin and doxorubicin for patients with Wilms' tumor: A report from the National Wilms' Tumor Study Group. *Journal of Clinical Oncology*, 16, 237-245.

Huang, Y. (2002), Calibration regression of censored lifetime medical cost. *Journal of the American Statistical Association*, 97, 318-327.

Jin, Z., Lin, D. Y., Wei, L. J., and Ying, Z. (2003), Rank-based inference for the accelerated failure time model. *Biometrika*, 90, 341-353.

Kalbfleisch, J. D. and Prentice, R. L. (2002), *The Statistical Analysis of Failure Time Data*, 2nd Ed., New York, John Wiley & Sons, Inc.

Kulich, M. and Lin, D. Y. (2000), Additive hazards regression for case-cohort studies, *Biometrika*, 87, 73-87.

Lin, D. Y., Wei, L. J., and Ying, Z. (1998), Accelerated failure time models for counting processes, *Biometrika*, 85, 605-618.

Little, R. J. A. and Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, 2nd Ed., Hoboken, New Jersey, John Wiley & Sons, Inc.

Nan, B., Emond, M., and Wellner, J. A. (2004), Information bounds for Cox regression models with missing data, *Annals of Statistics*, 32, 723-753.

Prentice, R. L. (1986), A case-cohort design for epidemiologic cohort studies and disease prevention trials, *Biometrika*, 73, 1-11.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (2002), *Numerical Recipes in C*, Cambridge University Press.

Ritov, Y. (1990), Estimation in a Linear Regression Model with Censored Data, *Annals of Statistics*, 18, 303-328.

Self, S. G. and Prentice, R. L. (1988), Asymptotic distribution theory and efficiency results for case-cohort studies, *Annals of Statistics*, 16, 64-81.

Tsiatis, A. A. (1990), Estimating regression parameters using linear rank tests for censored data, *Annals of Statistics*, 18, 354-372.

van der Vaart, A. W. and Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag New York, Inc.

Ying, Z. (1993), A large sample study of rank estimation for censored regression data, *Annals of Statistics*, 21, 76-99.

Table 1: Simulation summary statistics for estimating β in the model $\log T = \beta Z + \epsilon$ based on 500 replications with cohort size 5000.

Method	Mean of $(\hat{\beta})$	S.E. of $(\hat{\beta})$	Mean of $\hat{se}(\hat{\beta})$	90% CP	95% CP
		(1) $\beta = 0,\ p_z = 0.1$			
Full	0.003	0.095	0.098	91.2%	95.2%
SRS	-0.032	0.184	0.193	90.0%	94.8%
Strat0$^\sharp$	-0.011	0.193	0.200	89.4%	94.8%
Strat1†	-0.004	0.173	0.182	91.4%	96.6%
Strat2‡	0.007	0.143	0.151	90.4%	95.0%
		(2) $\beta = 1,\ p_z = 0.1$			
Full	1.007	0.161	0.179	94.2%	98.2%
SRS	0.979	0.228	0.275	94.0%	97.2%
Strat0$^\sharp$	0.990	0.237	0.273	93.8%	97.8%
Strat1†	0.996	0.214	0.254	95.8%	98.6%
Strat2‡	1.007	0.189	0.222	95.8%	98.2%
		(3) $\beta = 0,\ p_z = 0.3$			
Full	0.002	0.061	0.062	90.2%	94.8%
SRS	-0.011	0.130	0.119	87.8%	93.2%
Strat0$^\sharp$	-0.006	0.127	0.124	88.0%	92.4%
Strat1†	-0.004	0.117	0.118	89.6%	94.6%
Strat2‡	0.002	0.114	0.113	90.4%	94.8%
		(4) $\beta = 1,\ p_z = 0.3$			
Full	1.001	0.084	0.089	91.8%	96.0%
SRS	0.991	0.134	0.144	91.8%	95.8%
Strat0$^\sharp$	0.987	0.141	0.154	92.6%	97.8%
Strat1†	0.989	0.131	0.147	93.4%	97.4%
Strat2‡	0.993	0.128	0.140	92.8%	97.2%

$^\sharp$ Strat0 is stratified with $\eta = \nu = 0.5$, which is equivalent to non-stratification;
† Strat1 is stratified with $\eta = \nu = 0.7$;
‡ Strat2 is stratified with $\eta = \nu = 0.9$;

Table 2: Results for the Wilms tumor study based on 20 sampled subcohorts.

	Parameter estimates (S.E.) under different sampling							
	UH	Stage II	Stage III	Stage IV	NWTSG-4	log(Diam)	Age 2-4	Age $\geq$ 4
Full	-3.943	-1.502	-1.792	-3.097	0.053	-0.280	0.883	-0.326
	(0.175)	(0.391)	(0.377)	(0.281)	(0.209)	(0.314)	(0.354)	(0.239)
SRS†	-3.945	-1.481	-1.702	-3.205	-0.001	-0.483	0.967	-0.227
	(0.279)	(0.421)	(0.444)	(0.484)	(0.323)	(0.427)	(0.426)	(0.409)
Strat‡	-3.918	-1.564	-1.836	-3.303	0.052	-0.164	1.016	-0.344
	(0.261)	(0.429)	(0.445)	(0.499)	(0.326)	(0.433)	(0.416)	(0.414)

† SRS with 800 subjects in subcohort;
‡ Strat is stratified with about 360 subjects in the favorable local histology group, about 360 subjects in the unfavorable local histology group, and about 80 from the unknown local histology group.